**Supplementary Information Document**

***Multipoint calibration (Calibration curve)***

Most methods in analytical chemistry uses the linear function Eq. 1 as calibration technique to determine the amount of analyte in a sample. The least square method (LSM) is often used to estimate the slope (a) and the linear coefficient (b) of the calibration curve, Eqs 2 and 3, respectively [[[1]](#endnote-1)-[[2]](#endnote-2)[[3]](#endnote-3)[[4]](#endnote-4)].

According to this method (LSM), the most adequate calibration curve is the one that results the lowest value for the quadratic sum of the residues, which are obtained by the difference between the analytical signal *yi* and the expected analytical signal *ŷi* Eq. 4 for a set of “n” experimental points [4, [[5]](#endnote-5)].

 (1)

where:

*y* = instrumental response

*a* = slope

*x* = mass fraction

b = linear coefficient

 ⭢ ; 

(2)



(3)

where:

 = mass fraction

 = instrumental response i

*n* = number of measurements for calibration

 = mean instrumental response

 = mean mass fraction levels

 (4)

 = residue

 = instrumental response i

 = instrumental response predicted by the calibration equation

From the parameters “a” and “b”, it is possible to calculate the analyte mass fraction in the sample (x0) by the interpolation of the equation of the line obtained from the least squares method, according to Eq. 5.

 (5)

If a statistical model used to describe the calibration curve is adequate, then the residues represent only the random errors of the experimental measurements of *y*.

To confirm the absence of systematic errors, visual inspection of the residue graphic (y-axis) *versus* mass fraction levels (x-axis) is recommended, these graphs can provide the first indication of possible non-linearity, in addition to being able to highlight a heteroscedastic behavior of the data [5]. In these graphs, the residues must have constant variance and must be randomly distributed throughout the calibration range. If the residues increase or decrease proportionally with the increase in mass fraction, then the data are heteroscedastic and the use of the regression is not indicated. If the data present positive residues followed by negative residues, for example, then the calibration function may not be linear and the suitability for another mathematical model should be investigated [[[6]](#endnote-6)].

The EURACHEM/CITAC (1998) [[[7]](#endnote-7)] also suggests that in case of heteroscedasticity, the calibration data should be treated by the weighted least squares method. Additionally, Souza and Junqueira (2005) [1] reported that the LSM has the disadvantage of being very sensitive to the presence of outliers and Miller (1991) [6] described that even calibration curves with large random errors in the y-axis and evident curvature can present R2 values near 1.

Outlier tests are applied to detect and/or remove discrepant values [5]. The main tests used are of Dixon, Chauvenet and Grubbs.

In this work, it was utilized the Grubbs’ test for outliers. The Grubbs’ test, first checks the existence of an aberrant value (the minimum value and/or the maximum value of the data set) by comparing with value of Grubbs (Gcalc\_1; Eq.6) and the critical value (Gcrit), for a given level of significance. When Gcalc\_1>Gcrit the suspected value should be discarded [[[8]](#endnote-8)].

  (6)

where:

 = calculated value of Grubbs for a suspected value

 = suspected value

 = arithmetic mean of the experimental set

 = standard deviation of the experimental set

If a suspected value is discarded, then a new test must be performed for new data set, until there are no more discrepant values [8].

The evaluation of the normality of the residues of the calibration curve is another important test to be performed. The normality of the residues can be tested by the Shapiro-Wilk test Eq. 7 [[[9]](#endnote-9)].

  (7)



where:

 = calculated Shapiro-Wilk value

 = value of each (residue of the calibration curve) in ascending order

 = value of the residue n-i + 1

 = constants generated by means, variances and covariance of statistical order of a sample of size "n" for a Normal distribution [9].

According to this test, if Wcalc>Wstat for the established level of significance, then the residues of the calibration curve can be considered to come from a normal population [9].

The homoscedasticity of the curve was also investigated. The homoscedasticity test evaluates whether the variances of the curve residues are constant along the calibration range. This parameter can be evaluated by the Cochran’s test and applied when the number of observations is the same for all mass fraction levels Eq. 8 [[[10]](#endnote-10)].

 (8)

where:

Ccalc = calculated Cochran's value

 = value of the maximum variance selected from instrumental responses between mass fraction levels i

 = sum of the variances of the instrumental responses for each level of mass fraction i

The Ccalc value is then compared with the tabulated value Ctab for a given level of significance. If Ccalc < Ctab the homoscedasticity is confirmed and, therefore, there is no significant difference between the variances of the responses along the concentration range [10]. However, if the Cochran test indicates heteroscedasticity, the regression should be obtained by the weighted least squares method [1,10,[[11]](#endnote-11)].

Serial correlation of residues (autocorrelation) may cause underestimation of variance and confidence interval and may lead to erroneous inferences, such as indicating false significance of regression coefficients. Thus, assuming that residues are independent variables, all series correlations are equal to 0, that is, ρs = 0. The Durbin-Watson statistical test is used to test whether the null hypothesis ρs = 0 is true according to Eq. 9 [1]:

  (9)

where:

 = Durbin-Watson statistic

 = residue i

 = residue immediately prior to residue i

The value of Durbin-Watson varies from 0 to 4 with an average of 2. If the calculated value converges to 2, this means that there is no autocorrelation, on the other hand, if it approaches 0 or 4, autocorrelation increases. Thus, a value of d = 0 indicates perfect positive correlation while a value of d = 4 indicates perfect negative correlation. In practice, the test consists of comparing the value of “calculated d” (dcalc) with the lower limit (dL) and the upper limit (dU).

In order to test for positive autocorrelation at a given level of significance, the following criteria must be observed:

If d<dL - there is statistical evidence that the errors are positively autocorrelated.

If d>dU - there is no statistical evidence that the errors are positively autocorrelated.

If dL <d<dU - the test is inconclusive.

The limit values for dL and dU depend on the level of significance, on the number of residues and on the number of predictors [1]. The independence of the residues can be graphically illustrated by plotting each value of ei (x-axis) versus the respective value ei-1 (y-axis), which should indicate a random distribution [1].

The next parameters to be tested in the calibration curve are the lack of fit and the significance of the regression, which can be evaluated through the F test [[[12]](#endnote-12), [[13]](#endnote-13)]. In this test, the quadratic sum of the residues of the model (SSR) is decomposed into the quadratic sum corresponding to the pure error (SSPE) and quadratic sum corresponding to the lack of adjustment (SSLF) and they can be evaluated by Analysis of Variance (ANOVA).

Table 1 - Analysis of variance table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source of Variation** | **Sum of Squares (SS)** | **Degrees of Freedom (υ)** | **Mean Square (MQ)** | ***Fcalc*** |
| Regression (R) |  | *p-1* | *MSR = SSR/(p-1)* | *MSR/MSE* |
| Residual error (E) |  | *n-p* | *MSE = SSE/(n-p)* |
| Lack of fit (LF) |  | *k-p* | *MSLF = SSLF/(k-p)* | *MSLF/MSPE* |
| Pure error (PE) |  | *n-k* | *MSPE = SSPE/(n-k)* |
| **Total** |  | *n-1* | *---* | *---* |

= number of measurements at each mass fraction level, = predicted values for each measurement i, = mean value of all measured values,  = instrumental response for measurement i, = mean instrumental response at each mass fraction level j,  = number of parameters of the model, = number of measurements for the calibration curve, = number of mass fraction levels on the x-axis. SOURCE: [12, 13]

By dividing the quadratic sum of the pure error (SSPE) and the quadratic sum of the lack of adjustment (SSLF) by the respective degrees of freedom (υ), it is possible to obtain the quadratic mean due to the pure error (MSPE) and the quadratic mean due to the lack of fit (MSLF). It is expected that for a well-adjusted model, the quadratic mean due to lack of fit (MSLF) reflects only the random errors. Therefore, the MSLF value should not be significantly greater than the MSPE value.

To test if MSLF>MSPE, the MSLF/MSPE ratio is compared with the distribution value of F for the corresponding degrees of freedom (Ftab\_1). When the value of MSLF/MSPE<Ftab\_1 it is an indicative that the error due to the lack of fit is not significantly greater than the pure error and the model can be considered satisfactory [13].

Another parameter to be evaluated by the ANOVA test (Table 1) is the significance of the regression, made by comparing the ratio between the mean square of the regression (MSR) and the quadratic mean of the residue error (MSE) with the value of F for the respective degrees of freedom (Ftab\_2). Thus, it can be concluded that the greater the value of MSR/MSE, the better the significance of the regression, and if MSR/MSE>Ftab\_2 it can be concluded that the regression has statistical significance [13].

Thus, if the experimental data of the calibration curve are satisfactory, the linear adjustment obtained by the LSM can be used to determine the concentration of analyte in the sample.

**Sources of the uncertainty**

The uncertainty budget was carried out employing classic method according to ISO GUM guide.

***Multipoint calibration (Calibration curve)***

The measurement uncertainty was estimated from to specification of the measurand. The measurand is the mass fraction of SeMet in the yeast sample (*Saccharomyces cerevisiae*) and is defined by the Eq. 10.

 (10)

*dfn*

*moist*

Where:

*w(Se)* = mass fraction of SeMet in the yeast sample

*w(Se)0* = mass fraction of SeMet in the analytical solution

mtn = total mass associated to dilution(s) (1, 2, ..., n)

msn = sample mass associated to dilution(s) (1, 2, ..., n)

dfn = dilution fator(s)

frep = factor of the instrumental repeatability

*fext* = factor of the extraction procedure repeatability

*mi* = initial mass of the sample before water removal

*mf* = final mass of the sample after water removal

*fmoist* = factor of the moisture procedure repeatability

*moist* = moisture of the sample

From the specification of the measurand Eq. 10, the cause and effect diagram was constructed and the components that contribute to the measurement uncertainty include the uncertainty associated with the mass fraction of SeMet in the analytical solution (uw(Se)0), uncertainty associated with sample dilution (udfn), uncertainty associated with sample extraction (ufext), uncertainty associated with instrumental repeatability (ufrep) and uncertainty associated with moisture (umoist).



Fig.1. - Ishikawa (cause and effect) diagram of the uncertainties for the proposed method of fraction mass SeMet determination in the *Saccharomyces cerevisiae* by calibration curve

The SeMet mass fraction in the analytical solution (*w*(Se)0) was obtained by the linear regression using the least square method (LSM) and the uncertainty is considered of the Type A, this uncertainty source (*uw*(Se)0) was estimated according to the EURACHEM (11).



(11)

Where:

 = uncertainty associated to mass fraction of the SeMet in the analytical solution

 = standard deviation of the residues of the calibration curve

*a* = slope of the calibration curve

*p* = number of measures to determine *w(Se)0*.

*n* = number of instrumental replicates of the sample

 = Sum of the residual of standardmass fraction

 = mass fraction of the SeMet in the analytical solution

 = average value of the levels of mass fraction using in the calibration curve

The factor associated to the instrumental repeatability (*frep*) is considered uncertainty of Type A and had important contribution to the final uncertainty. This source of the uncertainty (*ufrep*) was estimated according to Eq. 12.

  Eq. 12

Where:

 = uncertainty associated to instrumental repeatability

 = standard deviation of the instrumental replicates

 = number of instrumental replicates

The factor associated to the moisture of the sample (*fmois*t) is also an uncertainty of Type A and was estimated according to Eq.13.

  (13)

Where:

 = uncertainty associated to variability of procedure for moisture determination

 = standard deviation of the moisture content

 = number of samples using to determine the moisture

The factor associated to the extraction procedure (fext) is an uncertainty of Type A and was estimated according to Eq. 14.

  (14)

Where:

 = uncertainty associated to extraction procedure

 = standard deviation between the averages of fraction masses of the samples submitted to the extraction procedure

 = number of samples submitted to the extraction procedure

The uncertainties associated with weighing procedures (mi, mf, mtn, msn) are uncertainties of type B. They are determined according to Eq 15.

  (15)

where:

 = uncertainty associated to mass

 = expanded uncertainty of the analytical balance

 = coverage factor of the analytical balance

After defining the measurand, identifying the sources of uncertainty, constructing the cause and effect diagram and estimating the uncertainty of the input quantities, the next step of estimating the measurement uncertainty by the classical method was to determine the sensitivity coefficients (*Ci*). These coefficients are used to describe how the input variable influences the output quantity and are calculated by partial derivative of the input variable *xi* with respect to the output quantity *y*, according to Eq. 16.



 (16)

Where:

ci = sensitivity coefficient of the variable i

 = partial derivative of the output quantity *y*

 = partial derivative of the input variable *xi*

So, the partial derivatives used in this work are present in the Table 1.

Table 1 - Partial derivative for estimative of the measurement uncertainty of the mass fraction of selenomethionine.

|  |  |  |
| --- | --- | --- |
| **Partial derivative** | **Estimative** | **Unit** |
|  |  | --- |
|  |  | mg kg-1 g-1 |
|  |  | mg kg-1 g-1 |
|  |  | --- |
|  |  | --- |
|  |  | mg kg-1 g-1 |
|  |  | mg kg-1 g-1 |
|  |  | mg kg-1 |

The uncertainty combined (*uc*) was estimated from all the uncertainty sources using the law of propagation of uncertainties. So, as all the input uncertainties are independent variables, the Eq. 17 was used to estimate the *uc*.

**(**17**)**

As described in ISO GUM, the uncertainties may be classified into two kinds: type A and type B. The first one is associated with statistical analysis of a series of observations, whereas Type B is approached by means other than the statistical analysis of a series of observations. This information is important in calculating the effective degrees of freedom Eq. 18 that is necessary to find the coverage factor (*k*) used in the expanded uncertainty U, Eq.19.

 (18)

Where:

 = effective degrees of freedom

= uncertainty sources of the Type A

= degrees of freedom of the respective uncertainty sources of the Type A

  (19)

where:

*k* = coverage factor for effective degrees of freedom (eff) at a 95% confidence level

*U* = the relative expanded uncertainty.

Finally, the measurement result of the measurand was declared as *w*(Se) ± U, complemented with information on probability and coverage factor k.

***Single point calibration***

The measurement uncertainty was estimated from to specification of the measurand. The measurand is the mass fraction of SeMet in the yeast sample (*Saccharomyces cerevisiae*) and is defined by the Eq.

 (20)

where:

*wSeMet*= mass fraction of SeMet in the sample

 = average of instrumental responses of SeMet in the analytical solution of the sample

 = mass of the sample

 = average of instrumental responses of SeMet in the analytical solution of the standard

 = mass of SeMet in the standard

*mi* = initial mass of the sample before water removal

*mf* = final mass of the sample after water removal

*fmoist* = factor of the moisture procedure repeatability

In this way, the cause and effect was constructed according to Fig. 2.



Fig. 2 - Ishikawa (cause and effect) diagram of the uncertainties for the proposed method of fraction mass SeMet determination in the *Saccharomyces cerevisiae* by single point calibration

The uncertainty sources *uysample*, u*yCRM,* u*moist* and u*rep*, should be considered as uncertainty of Type A and are calculated according to Eq. 21.

  (21)

where:

 = uncertainty associated to standard deviation

 = standard deviation

 = number of replicates

On the other hand, the uncertainty associated with *xMRC* (*uxMRC*) is calculated according to Eq. 22.

  (22)

Where:

 = uncertainty associated to mass fraction of SeMet of the standard

 = uncertainty expanded associated to certificate of the standard

 = uncertainty associated to mass

Once the sources of uncertainty for the *wSeMet* were estimated, the sensitivity coefficients (Table 2) were determined, based on the definition of the measurand, according to Eq. (16) and Eq. (20).

Table 2 - Partial derivative for estimative of the measurement uncertainty of the w(Se)0 by the one point calibration

|  |  |  |
| --- | --- | --- |
| **Partial derivatives** | **Estimative** | **Unit** |
|  |  | mg kg-1 cps-1 |
|  |  | 1 kg-2 |
|  |  | 1 kg-1 |
|  |  | mg kg-1 cps-1 |
|  |  | - |
|  |  | mg kg-2 |
|  |  | mg kg-2 |
|  |  | mg kg-1 |

So, the uncertainty associated with *wSe* was estimated according to Eq. 23.

  (23)

Finally, the next steps were to determine the sensitivity coefficient (*Ci*), the combined uncertainty (*uc*), the effective degrees of freedom () and the expanded uncertainty (*U*), which were determined according to Eqs 16, 23, 18, 19 respectively.

**References**

1. Souza, S. V. C.; Junqueira, R. G. *Anal. Chim. Acta* **2005**, 552, 25-35. [↑](#endnote-ref-1)
2. De Beer, J. O.; Naert, C. *Accred Qual Assur* **2012**,17, 265-274. [↑](#endnote-ref-2)
3. Nascimento, R. S.; Froes, R. E. S.; Silva, N. O. C.; Naveira, R. L. P.; Mendes, D. B. C.; Neto, W. B.; Silva, J. B. B. *Talanta* **2010**,80, 1102-1109. [↑](#endnote-ref-3)
4. Ribeiro, F. A. L.; Ferreira, M. M. C.; Morano, S. C.; Silva, L. R.; Schneider, R. P. *Química Nova* **2008**,31, 164-171. [↑](#endnote-ref-4)
5. Mulholland, M.; Hibbert, D. B. *J. Chromatogr. A* **1997**, 762, 73-82. [↑](#endnote-ref-5)
6. Miller, J. N.; *Analyst* **1991**, 116, 3-14. [↑](#endnote-ref-6)
7. EURACHEM/CITAC Guide; The Fitness for Purpose of Analytical Methods, EURACEM Working Group, December, **1998**. [↑](#endnote-ref-7)
8. OLIVEIRA, E. C.; Comparação das diferentes técnicas para a exclusão de “outliers”,

ENQUALAB, Congresso da Qualidade em Metrologia, 2008. [↑](#endnote-ref-8)
9. Shapiro, S. S.; Wilk, M. B. *Biometrika* **1965**, 52, 591-611. [↑](#endnote-ref-9)
10. González, A. G.; Herrador, M. A.; Asuero, A. G.; Sayago, A. *Accred Qual Assur* **2006**, 11, 256-258. [↑](#endnote-ref-10)
11. Mermet, J. *Spectrochim. Acta Part B* **2010**,65, 509-523. [↑](#endnote-ref-11)
12. The Pennsylvania State University Privacy and Legal Statements <https://onlinecourses.science.psu.edu/stat501/node/250>, accessed on November, 2017. [↑](#endnote-ref-12)
13. Pimentel, M. F; Neto, B. B. *Química Nova* **1996**, 19, 268 – 277. [↑](#endnote-ref-13)